International Journal of Computer Science Engineering
and Information Technology Research (IJCSEITR)
ISSN(P): 2249-6831; ISSN(E): 2249-7943
Vol. 4, Issue 5, Oct 2014, 9-12

SENTENCE LEVEL SENTIMENT CLASSIFICATION USING HMM WITH THE HELP OF PART OF SPEECH TAGGING

ANURAG MULKALWAR & KAVITA KELKAR

[1]Department of Computer Engineering, K. J. Somaiya College of Engineering, Vidya Vihar, Mumbai, Maharashtra, India

ABSTRACT

Sentiment analysis is a well known technique for finding sentiments from text data without any human intervention. Sentiment analysis shows various methods for implementation. This is the paper which suggests new approach towards classification of sentiments which are present in textual content. To support Hidden Markov Model it suggests some transition rules for model rather than transition probability. It effectively uses part of speech tagging for formation of transition rules.

KEYWORDS: Sentiment Analysis, Hidden Markov Model, POS Tagging

INTRODUCTION

Impact of Digital world on humans day-to-day life is quite impressive than past decade. Digital media now become daily need of person. There is massive growth in internet users noted in past five to six year and this number is still increasing. Surprising thing is that internet users are not only belongs to single age group or civilization but it is spread in all levels of civilization. By taking this advantageous situation as a future scope numerous web based applications, e-businesses start launching themselves on this huge but comfortable platform. Best web based applications are Social networking websites like facebook, twitter. And businesses are like Movies, Products promotion websites. Still all this businesses are very good at their place but it is also important thing needs to be considered that user response to this things. Internet users leave their response in the form of reviews, comments, blogs etc. It is very time consuming task to go throw each and every review and to understand whether user talking positively or negatively. To Solve this complex situation a very effective process come in to consideration that is Sentiment analysis. Sentiment analysis is possible at various levels i.e. Aspect level, Sentence level, Document level. At Aspect level of classification sentiments are classified according to features. At Document level of classification sentiments from whole documents are get considered and document get classified to its sentiment class. At sentence level of classification, sentiments from sentence taking into account to classify that sentence into its sentiment class. This paper will talk about a new approach toward classification of sentiment but not at document level but it is at sentence level. In this paper basis classification technique used is Hidden Markov Model classifier. Here we are going to discus all the modifications which are made for exact sentiment classification.

METHODOLOGY

Refining Textual Content

Before taking whole document as it is for classification it is a mandatory task to convert this document in to a acceptable processing unit. Without doing this task classification process may not be that much easy. To build actual processing unit we need to perform some operations on document. This operations are sentence separation, tokenization,

removal of useless words from sentence, assign POS value to each token, assign polarity to that tokens (positive or negative).

Sentence Separation: We firstly divide whole document into individual sentences. While parsing through document using full stop as a criteria of single sentence all sentences are get separated.

Tokenization: For tokenization we consider one sentence at a time and take it as input string. All tokens which are formed after tokenization always associated with that sentence only. Tokenization process remove punctuation marks and after occurrence of white space current word consider as token.

Stop-Word Removal: Stop words are such words from sentence which takes only part in sentence completion and that words are repeated continuously throughout the documents. Presence of such words increase computational cost. To avoid extra computation cost we exclude this stop words from tokens.

Part of Speech(POS) Tagging

Sentence from any language get its exact structure because of part of speech. English language there are mainly four categories of part of speech. Those are nouns, verbs, adverbs, and adjectives. Each part of speech has its own role in sentence formation. For our classification process its required to find POS of each token. There are various open source programs are available for part of speech tagging. Table 1 shows POS notations

Table 1: Part of Speech (POS) Notations

Noun	N
Verb	V
Adjective	J
Adverb	R

Sentiment Class Notation(SCN) Tagging

After tagging POS next task is to assign sentiment notation to the token, because each token has their desired sentiment class i.e. positive (Ps) or negative(Ng). When we observe sentence carefully we find that most of the nouns and verbs doesn't contain any sentiment. In this process we avoid such nouns and verbs from sentiment notation tagging.

To assign proper sentiment class notation to token there will be necessity of pre-classified tokens cluster based on POS and its sentiment class. Figure 1 shows cluster of POS.

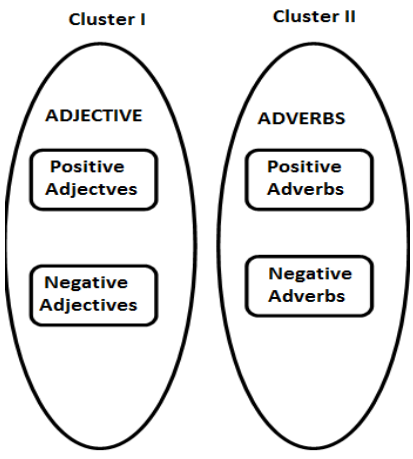

Figure1: POS Cluster Using Sentiment Class

Hidden Markov Model

Hidden Markov model is probabilistic model. It uses transition probabilities of observation associated with class for transition from one class to another class.HMM depends on five variables. M= (I, E, T, O, S).

In above vector S is a set of Possible States of classification S={Positive, Negative}. O is set of Observations O={Set of occurrence of tokens}. T is transition probabilities. I is initial probabilities of class. Figure 2 shows classification of sequence of observations.

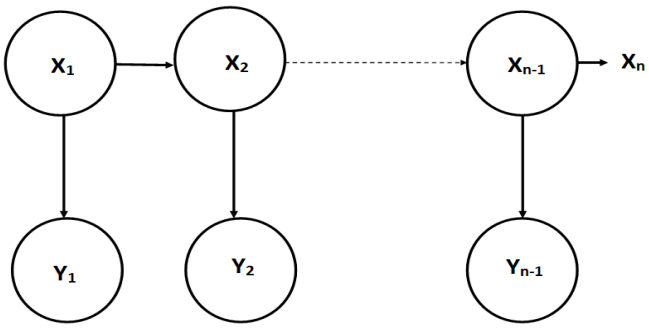

Figure 2: Hidden Markov Model Classifier

In this figure 2 X_1, X_2,……. X_n class labels from class set S, and Y_1, Y_2,……. Y_n are sequence of observations.

$$P(X_1| Y_1)=P(Y_1)P(Y_1| X_1) \quad (1)$$

$$P(X_n| Y_{1:\,n-1})=\sum_{n-1} P(X_n |X_{n-1})\, P(X_{n-1}|Y_{1:n-1}) \quad (2)$$

Equation (2) gives predicted class of series of observations. In our classification we are not using transition probabilities instead of that we are going to use transition rules. Transition rules states that which observation sequence will leads to transition from one class to another class. To design transition rule we use tokens with their assigned POS tag and SCN tag. For example suppose “good” is a token and it will be after tagging “good_J_Ps”. It stat that good is a adjective with positive sentiment. Table 2 shows transition rules.

Table 2: Transition Rules

Previous Tag	Current Tag	Transition Class
J_Ps	J_Ps	Positive
J_Ng	J_Ng	Negative
R_Ps	J_Ps	Positive
J_Ps	R_Ps	Positive
R_Ng	J_Ng	positive
R_Ng	J_Ps	negative

Based on above transition rules and their previous class HMM predicts its next class.

CONCLUSIONS

In this paper we proposed Text classification method using Hidden Markov Model classifier with the help of part of speech(POS) tagging and Sentiment class Notation(SCN) tagging. This process will give best result when domain knowledge is strong. Based on domain knowledge better construction of POS cluster is possible. Whole approach is based

on POS cluster and Transition rule. If those part get written strongly then this model will be the best Sentiment text classifier model.

REFERENCES

1. Sowmya Kamath S, Anusha Bagalkotkar, Ashesh Khandelwal, Shivam Pandey, Kumari Poornima. (2013). "Sentiment Analysis Based Approaches for Understanding User Context in Web Content", International Conference on Communication Systems and Network Technologies
2. Bing Liu. (2012). Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers.
3. Keke Cai*, Scott Spangler!, Ying Chen!, Li Zhang. (2008). **"**Leveraging Sentiment Analysis for Topic Detection" IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology ,
4. Bo Pang and Lillian Lee, Shivakumar Vaithyanathan. (July 2002)."Thumbs up? Sentiment Classification using Machine Learning Techniques", Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia.
5. FAN Na , CAI Wan-dong, ZHAO Yu. (2009). "A Method based on Generation Models for Analyzing Sentiment-Topic in Texts" IEEE.
6. Si Li, Hao Zhang, Weiran Xu, Guang Chen and Jun Guo. (2010). " Exploiting Combined Multi-level Model for Document Sentiment Analysis", International Conference on Pattern Recognition.
7. S.M.Shamimul Hasan, Donald A. Adjeroh. (2011). "Proximity-Based Sentiment Analysis", IEEE.
8. Lizhen Liu, Xinhui Nie, Hanshi Wang. (2012). "Toward a Fuzzy Domain Sentiment Ontology Tree for Sentiment Analysis" 5th International Congress on Image and Signal Processing (CISP).
9. Zhen Niu, Zelong Yin, Xiangyu Kong. (2012). "Sentiment Classification for Microblog by Machine Learning", Fourth International Conference on Computational and Information Sciences.
10. Mostafa Karamibekr, Ali A. Ghorbani. (2012). "Sentiment Analysis of Social Issues", International Conference on Social Informatics.
11. Samatcha Thanangthanakij, Eakasit Pacharawongsakda, Nattapong Tongtep, Pakinee immanee, Thanaruk Theeramunkong. (2012). "An Empirical Study on Multi-Dimensional Sentiment Analysis from User Service Reviews", Seventh International Conference on Knowledge, Information and Creativity Support Systems.
12. Peter Koncz and Jan Paralic. (2011). "An approach to feature selection for sentiment analysis", 15th International Conference on Intelligent Engineering Systems.
13. Wang Zuhui Jiang Wei. (2012). "Online Reviews Sentiment Analysis Applying Mutual Information", 9th International Conference on Fuzzy Systems and Knowledge Discovery.